



Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2007-060
CBCL-269

December 26, 2007

Learning complex cell invariance from natural videos: A plausibility proof

Timothee Masquelier, Thomas Serre, Simon
Thorpe, and Tomaso Poggio

Report Documentation Page		Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE 26 DEC 2007	2. REPORT TYPE	3. DATES COVERED 00-00-2007 to 00-00-2007
4. TITLE AND SUBTITLE Learning complex cell invariance from natural videos: A plausibility proof		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER	
	5e. TASK NUMBER	
	5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory ,32 Vassar Street, Cambridge, MA, 02139		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited		
13. SUPPLEMENTARY NOTES		

14. ABSTRACT

One of the most striking feature of the cortex is its ability to wire itself. Understanding how the visual cortex wires up through development and how visual experience refines connections into adulthood is a key question for Neuroscience. While computational models of the visual cortex are becoming increasingly detailed (Riesenhuber and Poggio, 1999; Giese and Poggio, 2003; Deco and Rolls, 2004, 2005; Serre et al., 2005a; Rolls and Stringer, 2006; Berzhanskaya et al., 2007; Masquelier and Thorpe, 2007), the question of how such architecture could self-organize through visual experience is often overlooked. Here we focus on the class of hierarchical feedforward models of the ventral stream of the visual cortex (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Mel, 1997; VanRullen et al., 1998; Riesenhuber and Poggio, 1999; Ullman et al., 2002; Hochstein and Ahissar, 2002; Amit and Mascaró, 2003; Wersing and Koerner, 2003; Serre et al., 2005a; Masquelier and Thorpe, 2007; Serre et al., 2007), which extend the classical simple-to-complex cells model by Hubel and Wiesel (1962) to extra-striate areas, and have been shown to account for a host of experimental data. Such models assume two functional classes of simple and complex cells with specific predictions about their respective wiring and resulting functionalities. In these networks, the issue of learning, especially for complex cells, is perhaps the least well understood. In fact, in most of these models, the connectivity between simple and complex cells is not learned but rather hard-wired. Several algorithms have been proposed for learning invariances at the complex cell level based on a trace rule to exploit the temporal continuity of sequences of natural images (e.g., (Földiák 1991; Wallis and Rolls, 1997; Wiskott and Sejnowski, 2002; Einh user et al., 2002; Spratling, 2005)), but very few can learn from natural cluttered image sequences. Here we propose a new variant of the trace rule that only reinforces the synapses between the most active cells, and therefore can handle cluttered environments. The algorithm has so far been developed and tested at the level of V1-like simple and complex cells: we verified that Gabor-like simple cell selectivity could emerge from competitive Hebbian learning (see also (Delorme et al., 2001; Einh user et al., 2002; Guyonneau, 2006)). In addition, we show how the modified trace rule allows the subsequent complex cells to learn to selectively pool over simple cells with the same preferred orientation but slightly different positions thus increasing their tolerance to the precise position of the stimulus within their receptive fields.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT

unclassified

b. ABSTRACT

unclassified

c. THIS PAGE

unclassified17. LIMITATION OF
ABSTRACT**Same as
Report (SAR)**18. NUMBER
OF PAGES**20**19a. NAME OF
RESPONSIBLE PERSON

Learning complex cell invariance from natural videos: a plausibility proof

Timothée Masquelier, Thomas Serre, Simon Thorpe and Tomaso Poggio

Abstract

One of the most striking feature of the cortex is its ability to wire itself. Understanding how the visual cortex wires up through development and how visual experience refines connections into adulthood is a key question for Neuroscience. While computational models of the visual cortex are becoming increasingly detailed (Riesenhuber and Poggio, 1999; Giese and Poggio, 2003; Deco and Rolls, 2004, 2005; Serre et al., 2005a; Rolls and Stringer, 2006; Berzhanskaya et al., 2007; Masquelier and Thorpe, 2007), the question of how such architecture could self-organize through visual experience is often overlooked.

Here we focus on the class of hierarchical feedforward models of the ventral stream of the visual cortex (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Mel, 1997; VanRullen et al., 1998; Riesenhuber and Poggio, 1999; Ullman et al., 2002; Hochstein and Ahissar, 2002; Amit and Mascaró, 2003; Wersing and Koerner, 2003; Serre et al., 2005a; Masquelier and Thorpe, 2007; Serre et al., 2007), which extend the classical simple-to-complex cells model by Hubel and Wiesel (1962) to extra-striate areas, and have been shown to account for a host of experimental data. Such models assume two functional classes of simple and complex cells with specific predictions about their respective wiring and resulting functionalities.

In these networks, the issue of learning, especially for complex cells, is perhaps the least well understood. In fact, in most of these models, the connectivity between simple and complex cells is not learned but rather hard-wired. Several algorithms have been proposed for learning invariances at the complex cell level based on a trace rule to exploit the temporal continuity of sequences of natural images (*e.g.*, (Földiák, 1991; Wallis and Rolls, 1997; Wiskott and Sejnowski, 2002; Einhäuser et al., 2002; Spratling, 2005)), but very few can learn from natural cluttered image sequences.

Here we propose a new variant of the trace rule that only reinforces the synapses between the most active cells, and therefore can handle cluttered environments. The algorithm has so far been developed and tested at the level of V1-like simple and complex cells: we verified that Gabor-like simple cell selectivity could emerge from competitive Hebbian learning (see also (Delorme et al., 2001; Einhäuser et al., 2002; Guyonneau, 2006)). In addition, we show how the modified trace rule allows the subsequent complex cells to learn to selectively pool over simple cells with the same preferred orientation but slightly different positions thus increasing their tolerance to the precise position of the stimulus within their receptive fields.

Copyright ©Massachusetts Institute of Technology, 2007

This report describes research done at the Center for Biological & Computational Learning, which is in the McGovern Institute for Brain Research at MIT, as well as in the Dept. of Brain & Cognitive Sciences, and which is affiliated with the Computer Sciences & Artificial Intelligence Laboratory (CSAIL).

This research was sponsored by grants from: DARPA Contract No. HR0011-04-1-0037, DARPA Contract No. FA8650-06-7632, National Science Foundation-NIH (CRCNS) Contract No. EIA-0218506, and National Institutes of Health (Conte) Contract No. 1 P20 MH66239-01A1. Additional support was provided by: Central Research Institute of Electric Power Industry (CRIEPI), Daimler-Chrysler AG, Eastman Kodak Company, Honda Research Institute USA, Inc., Komatsu Ltd., Merrill-Lynch, NEC Fund, Oxygen, Siemens Corporate Research, Inc., Sony, Sumitomo Metal Industries, Toyota Motor Corporation, and the Eugene McDermott Foundation.

1 Introduction

Learning is arguably the key to understanding intelligence (Poggio and Smale, 2003). One of the most striking feature of the cortex is its ability to wire itself. Understanding how the visual cortex wires up through development and how plasticity refines connections into adulthood is likely to give necessary constraints to computational models of visual processing. Surprisingly there have been relatively few computational studies (Perrett et al., 1984; Földiák, 1991; Hietanen et al., 1992; Wallis et al., 1993; Wachsmuth et al., 1994; Wallis and Rolls, 1997; Stringer and Rolls, 2000; Rolls and Milward, 2000; Wiskott and Sejnowski, 2002; Einhäuser et al., 2002; Spratling, 2005) that have tried to address the mechanisms by which learning and plasticity may shape the receptive fields (RFs) of neurons in the visual cortex.

Here we study biologically plausible mechanisms for the learning of both selectivity and invariance of cells in the primary visual cortex (V1). We focus on a specific class of models of the ventral stream of the visual cortex, the feedforward hierarchical models of visual processing (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Mel, 1997; VanRullen et al., 1998; Riesenhuber and Poggio, 1999; Ullman et al., 2002; Hochstein and Ahissar, 2002; Amit and Mascaró, 2003; Wersing and Koerner, 2003; Serre et al., 2005a; Masquelier and Thorpe, 2007; Serre et al., 2007), which extend the classical simple-to-complex cells model by Hubel and Wiesel (1962) (see Box 1) and have been shown to account for a host of experimental data.

We have used a specific implementation of such feedforward hierarchical models (Riesenhuber and Poggio, 1999; Serre et al., 2005b, 2007), which makes predictions about the nature of the computations and the specific wiring of simple and complex units, denoted S_1 and C_1 units respectively. Learning in higher stages of the model will be addressed in future work. We show that with simple biologically plausible learning rules, these two classes of cells can be learned from natural real-world videos with no supervision. In particular, we verified that the Gabor-like selectivity of S_1 units could emerge from competitive Hebbian learning (see also (Delorme et al., 2001; Einhäuser et al., 2002; Guyonnet, 2006)). In addition, we proposed a new mechanism, which suggests how the specific pooling from S_1 to C_1 unit could self-organize by passive exposure to natural input video sequences. We discuss the computational requirements for such unsupervised learning to take place and make specific experimental predictions.

1.1 Evidence for learning and plasticity in the visual cortex

In the developing animal, ‘rewiring’ experiments (see (Horng and Sur, 2006) for a recent review), which reroute inputs from one sensory modality to an area nor-

mally processing a different modality, have now established that visual experience can have a pronounced impact on the shaping of cortical networks. How plastic is the adult visual cortex is however still a matter of debates.

From the computational perspective, it is very likely that learning may occur in all stages of the visual cortex. For instance if learning a new task involves high-level object-based representations, learning is likely to occur high-up in the hierarchy, at the level of IT or PFC. Conversely, if the task to be learned involves the fine discrimination of orientations like in perceptual learning tasks, changes are more likely to occur in lower areas at the level of V1, V2 or V4 (see (Ghose, 2004) for a review). It is also very likely that changes in higher cortical areas should occur at faster time scales than changes in lower areas.

By now there has been several reports of plasticity in all levels of the ventral stream of the visual cortex (see (Kourtzi and DiCarlo, 2006), *i.e.*, both in higher areas like PFC (Rainer and Miller, 2000; Freedman et al., 2003; Pasupathy and Miller, 2005) and IT (see for instance (Logothetis et al., 1995; Rolls, 1995; Kobatake et al., 1998; Booth and Rolls, 1998; Erickson et al., 2000; Sigala and Logothetis, 2002; Baker et al., 2002; Jagadeesh et al., 2001; Freedman et al., 2006) in monkeys or the LOC in humans (Dolan et al., 1997; Gauthier et al., 1999; Kourtzi et al., 2005; Op de Beeck et al., 2006; Jiang et al., 2007). Plasticity has also been reported in intermediate areas like in V4 (Yang and Maunsell, 2004; Rainer et al., 2004) or even lower areas like V1 (Singer et al., 1982; Karni and Sagi, 1991; Yao and Dan, 2001; Schuett et al., 2001; Crist et al., 2001), although their extent and functional significance is still under debate (Schoups et al., 2001; Ghose et al., 2002; DeAngelis et al., 1995).

At the cellular level, supervised learning procedures to validate Hebb’s covariance hypothesis *in vivo* have also been proposed. The covariance hypothesis predicts that a cell’s relative preference between two stimuli could be displaced towards one of them by pairing its presentation with imposed increased responsiveness (through iontophoresis). Indeed it was shown possible to durably change some cells’ RF properties in cat primary visual cortex, such as ocular dominance, orientation preference, interocular orientation disparity and ON or OFF dominance, both during the critical developmental period (Frégnac et al., 1988) and in adulthood (McLean and Palmer, 1998; Frégnac and Shulz, 1999). More recently, a similar procedure was used to validate the Spike Timing Dependent Plasticity in developing rat visual cortex (Meliza and Dan, 2006) using *in vivo* whole-cell recording.

Altogether the evidence suggests that learning plays a key role in determining the wiring and the synaptic weights of cells in the visual cortex.

Box 1: The Hubel & Wiesel hierarchical model of primary visual cortex.

Following their work on striate cortex, Hubel & Wiesel described a hierarchy of cells in the primary visual cortex: At the bottom of the hierarchy, the *radially symmetric* cells are like LGN cells and respond best to small spots of light. Second, the *simple* cells do not respond well to spots of light and require bar-like (or edge-like) stimuli at a particular orientation, position and phase (i.e., white bar on a black background or dark bar on a white background). In turn, the *complex* cells are also selective for bars at a particular orientation but they are insensitive to both the location and the phase of the bar within their receptive fields (RFs). At the top of the hierarchy the *hypercomplex* cells not only respond to bars in a position and phase invariant way, just like complex cells, but are also selective for bars of a particular length (beyond a certain length their response starts decreasing).

Hubel & Wiesel suggested that such increasingly complex and invariant object representations could be progressively built by integrating convergent inputs from lower levels. For instance, as illustrated in Fig. 1 (reproduced from (Hubel and Wiesel, 1962)), position invariance at the complex cells level, could be obtained by pooling over simple cells at the same preferred orientation but at slightly different positions.

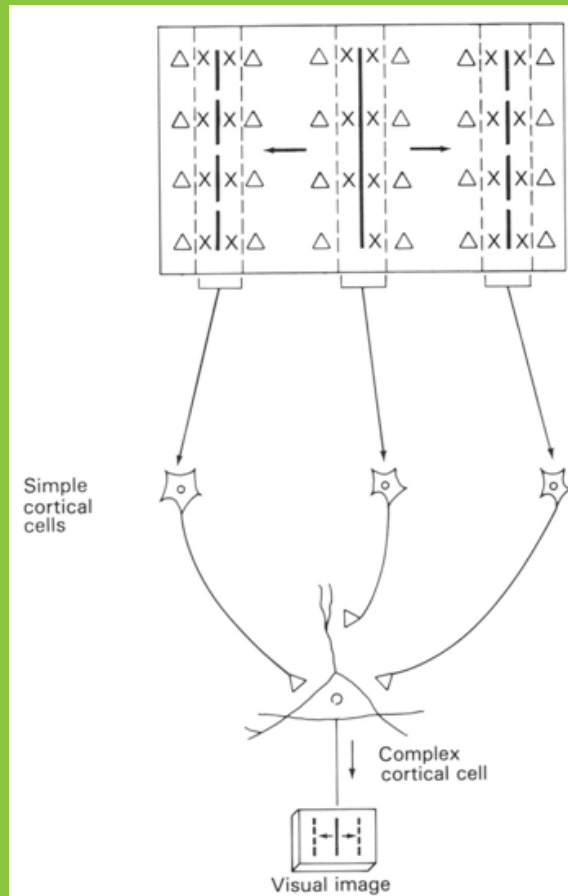


Figure 1: The Hubel & Wiesel hierarchical model for building complex cells from simple cells. Reproduced from (Hubel and Wiesel, 1962).

1.2 Simple and complex cell modeling

The key computational issue in object recognition is the specificity-invariance trade-off: recognition must be able to finely discriminate between different objects or object classes while at the same time be tolerant to object transformations such as scaling, translation, illumination, changes in viewpoint, changes in clutter, as well as non-rigid transformations (such as a change of facial expression) and, for the case of categorization, also to variations in shape within a class. Thus the main computational difficulty of object recognition is achieving a trade-off between selectivity and invariance. Extending the hierarchical model by (Hubel and Wiesel, 1962) (see Box 1) to extrastriate areas and based on theoretical considerations, Riesenhuber and Poggio (1999) speculated that only two functional classes of units may be necessary to achieve this trade-off:

The simple S units perform a TUNING operation over their afferents to build object-selectivity. The simple S units receive convergent inputs from retinotopically

organized units tuned to *different preferred stimuli* and combine these *subunits* with a bell-shaped tuning function, thus increasing object selectivity and the complexity of the preferred stimulus (see (Serre et al., 2005a) for details).

The analog of the TUNING operation in computer vision is the *template matching* operation between an input image and a stored representation. As discussed in (Poggio and Bizzzi, 2004) neurons with a Gaussian-like bell-shape tuning are prevalent across cortex. For instance simple cells in V1 exhibit a Gaussian tuning around their preferred orientation (Hubel and Wiesel, 1962) or even cells in inferotemporal cortex are typically tuned around a particular view of their preferred object (Logothetis et al., 1995; Booth and Rolls, 1998). From the computational point of view, Gaussian-like tuning profiles may be key in the generalization ability of cortex and networks that combine the activity of several units tuned with a Gaussian profile to different training examples have proved to be powerful learning scheme (Poggio and Girosi, 1990; Poggio and Smale, 2003).

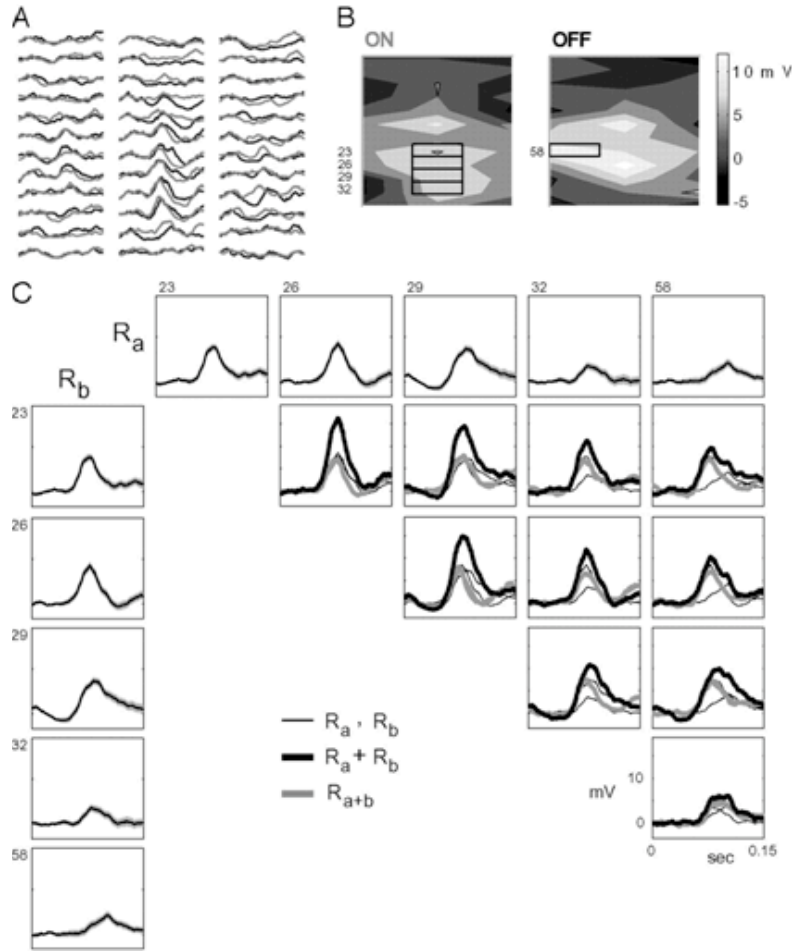


Figure 2: MAX-like operation from a complex cell in area 17 of the cat. Illustrated is the response of a complex cell to the simultaneous presentation of two bars (see (Lampl et al., 2004)) for details). A: average membrane potential measured from the response of the cell to bars of the optimal orientation. Black traces are the responses to dark bars (OFF responses) and gray traces are the responses to bright bars (ON responses). B: intensity plots obtained from the mean potentials. C: cell responses to each of the selected bars shown in B by thick lines around the rectangles. Lines in the 1st row and 1st column panels are the averaged responses to the presentation of a single bar, and the shaded area shows the mean ($\pm SE$). The inner panels present the response of the cell to the simultaneous presentation of the 2 bars whose positions are given by the corresponding column and row (gray traces), the responses to the 2 stimuli presented individually (thin black traces) and the linear sum of the 2 individual responses (thick black traces). Modified from (Lampl et al., 2004).

The complex C units receive convergent inputs from retinotopically organized S units tuned to the *same preferred stimuli* but at slightly different positions and scales with a MAX-like operation, thereby introducing tolerance to scale and translation. MAX functions are commonly used in signal processing (e.g., for selecting peak correlations) to filter noise out. The existence of a MAX operation in visual cortex was predicted by (Riesenhuber and Poggio, 1999) from theoretical arguments (and limited experimental evidence (Sato, 1989)) and was later supported experimentally in V4 (Gawne and Martin, 2002) and in V1 at the complex cell level (Lampl et al., 2004). Note that a soft-max close to an average may be sufficient for robust and invariant object recognition (Serre and Poggio, 2005) and seems to account for a significant proportion of complex cells (Finn

and Ferster, 2007). Fig. 2 (reproduced from (Lampl et al., 2004)) illustrates how a complex cell may combine the response of oriented retinotopically organized subunits (presumably simple cells) at the same preferred orientation with a MAX pooling mechanism.

Computational implementation of the Hubel & Wiesel model: In this work we use static idealized approximation to describe the response of simple and complex units. As described in (Knoblich et al., 2007; Kouh and Poggio, 2007; Serre et al., 2005a) both operations can be carried out by a *divisive normalization* followed by *weighted sum* and rectification. Normalization mechanisms (also commonly referred to as *gain control*) in this case, can be achieved by a feedforward (or recurrent) shunting inhibition (Torre and Poggio, 1978; Reichardt et al., 1983; Carandini and Heeger, 1994).²

Box 2: Computational implementation of the Hubel & Wiesel model.

We denote by (x_1, x_2, \dots, x_n) the set of inputs to a unit and $\mathbf{w} = (w_1, \dots, w_N)$ their respective input strength. For a complex unit, the inputs x_j are retinotopically organized and selected from an $m \times m$ grid of afferent units¹ Of afferents with the same selectivity (e.g., for an horizontal complex cells, subunits are all tuned to an horizontal bar but at slightly different positions and spatial frequencies). For a simple unit, the subunits are also retinotopically organized (selected from an $m \times m$ grid of possible afferents). But, in contrast with complex units, the subunits of a simple cell could in principal be with different selectivities to increase the complexity of the preferred stimulus. Mathematically, both the TUNING operation and the MAX operation at the simple and complex units level can be well approximated by the following equation:

$$y = \frac{\sum_{j=1}^n w_j^* x_j^p}{k + \left(\sum_{j=1}^n x_j^q \right)^r},$$

where y is the output of the unit, $k < 1$ is a constant to avoid zero-divisions and p, q and r represent the static non-linearities in the underlying neural circuit.

Such non-linearity may correspond to different regimes on the $f - I$ curve of the presynaptic neurons such that different operating ranges provide different degrees of non-linearities (from near-linearity to steep non-linearity). An extra sigmoid transfer function on the output $g(y) = 1/(1 + \exp^{\alpha(y-\beta)})$ controls the sharpness of the unit response.

By adjusting these non-linearities, the equation above can approximate better a MAX or a TUNING function:

- When $p \lesssim qr$, the unit approximates a Gaussian-like TUNING, *i.e.*, its response y will have a peak around some value proportional to the input vector $\mathbf{w} = (w_1, \dots, w_N)$. For instance, when $p = 1, q = 2$ and $r = 1/2$, the circuits perform a normalized dot-product with an L_2 norm, which with the addition of a bias term may approximate a Gaussian function very closely (see (Kouh and Poggio, 2007; Serre et al., 2005a) for details).
- When $p \gtrsim q + 1$ ($w_j \approx 1$), the unit implements a soft-max and approximates a MAX function very closely for larger q values (see (Yu et al., 2002), the quality of the approximation also increases as the inputs become more dissimilar). For instance, $r \approx 1, p \approx 1, q \approx 2$ gives a good approximation of the MAX (see (Kouh and Poggio, 2007; Serre et al., 2005a) for details).

The detailed mathematical formulation of the two operations is given in Box 2. There are plausible local circuits (Serre et al., 2005a) implementing the two key operations within the time constraints of the experimental data (Perrett et al., 1992; Keysers et al., 2001; Hung et al., 2005) based on small local population of spiking neurons firing probabilistically in proportion to the underlying analog value (Smith and Lewicki, 2006) and on shunting inhibition (Grossberg, 1973). Other possibilities may involve spike timing in individual neurons (Masquelier and Thorpe, 2007) (see (VanRullen et al., 2005) for a recent review). A complete description of the two operations, a summary of the evidence as well as plausible biophysical circuits to implement them can be found in (Knoblich et al., 2007; Serre et al., 2005a).

While there exists at least partial evidence for the existence of both Gaussian TUNING and max-like operations (see earlier), the question of how the specific wiring of simple and complex cells could self-organize during development and how their selectivity could be shape through visual experience is open. In the next

section, we review related work and speculate on computational mechanisms that could underlie the development of such circuits.

1.3 On learning simple and complex cells

Here we speculate that correlations play a key role in learning. Beyond the Hebbian doctrine, which says that ‘neurons that fire together wire together’, we suggest that correlation in the inputs of neurons could explain the wiring of both simple and complex cells. As emphasized by several authors, statistical regularities in natural visual scenes may provide critical cues to the visual system to solve specific tasks (Richards et al., 1992; Knill and Richards, 1996; Callaway, 1998; Coppola et al., 1998) or even provide a teaching signal (Barlow, 1961; Sutton and Barto, 1981; Földiák, 1991) for learning with no supervision. More specifically, we suggest that the wiring of the simple S units depends on learning correlations in *space* while the wiring of the C units depends on learning correlations in *time* (Serre et al., 2005a).

The wiring of the simple units corresponds to learning correlation between inputs **at the same time** (*i.e.*, for simple S_1 units in V1, the bar-like arrangements of LGN inputs, and beyond V1, more elaborate arrangements of bar-like subunits, *etc.*). This corresponds to learning which combinations of features appear most frequently in images. That is, a simple unit has to detect *conjunctions* of inputs (*i.e.*, sets of inputs that are consistently co-active), and to become selective to these patterns. This is roughly equivalent to learning a dictionary of image patterns that appear with higher probability.

This is a very simple and natural assumption. Indeed it follows a long tradition of researchers that have suggested that the visual system, through visual experience and evolution, may be adapted to the statistics of its natural environment (Attneave, 1954; Barlow, 1961; Atick, 1992; Ruderman, 1994) (see also (Simoncelli and Olshausen, 2001) for a review). For instance, Attneave (1954) proposed that the goal of the visual system is to build an efficient representation of the visual world and (Barlow, 1961) emphasized that neurons in cortex try to reduce the redundancy present in the natural environment.

This type of learning can be done with an Hebbian learning rule (von der Malsburg, 1973; Földiák, 1990). Here we used a slightly modified Hebb rule, which has the advantage of keeping the synaptic weights bounded, while remaining a local learning rule (see Section 4, Eq. 5). At the same time, a mechanism is necessary to prevent all the simple units in a given cortical column from learning the same pattern. Here we used hard competition of the 1-Winner-Take-All form (see (Rolls and Deco, 2002) for evidence). In the algorithm we describe below, at each iteration and within each hypercolumn only the most activated unit is allowed to learn (but it will do so if and only if its activity is above a threshold, see Section 4.3). In the cortex such a mechanism could be implemented by short range lateral inhibition.

Networks with anti-Hebbian horizontal connections have also been proposed (Földiák, 1990). While such networks could, in principle, remove redundancy more efficiently, horizontal connections are unlikely to play a critical role during the initial feedforward response of neurons within the first 10-30 ms after response onset (Thorpe and Imbert, 1989; Thorpe and Fabre-Thorpe, 2001; Keyser et al., 2001; Rolls, 2004). They could nevertheless be easily added in future work. Furthermore, a certain level of redundancy is desirable, to handle noise and loss of neurons.

Matching pursuit, which could also be implemented in the visual cortex via horizontal connections, has been proposed to reduce the redundancy and increases the sparseness of neuronal responses (Perrinet et al., 2004).

Previous work has already shown how selectivity to orientation could emerge naturally with simple

learning rules like Spike-Timing-Dependant-Plasticity (STDP) (Delorme et al., 2001; Guyonneau, 2006) and a Hebbian rule (Einhäuser et al., 2002). The goal of the work here is to apply such rule with a specific implementation of a model of the ventral stream of the visual cortex (Riesenhuber and Poggio, 1999; Serre et al., 2005a, 2007), formerly known as HMAX.

The wiring of complex units, on the other hand, may reflect learning from visual experience how to associate frequent transformations in time – such as translation and scale – of specific image features coded by simple cells. The wiring of the C units reflects learning of correlations **across time**, *e.g.*, for complex C_1 units, learning which afferent S_1 units with the same orientation and neighboring locations should be wired together because, often, such a pattern changes smoothly in time (under translation) (Földiák, 1991; Wiskott and Sejnowski, 2002).

As discussed earlier, the goal of the complex units is to increase the invariance of the representation along one stimulus dimension. This is done by combining the activity of a group of neighboring simple units tuned to the same preferred stimulus at slightly different positions and scales. In this work we focus on translation invariance, but the same mechanism, in principle, could be applied to any transformation (for *e.g.*, scale, rotation or view-point).

A key question is how a complex cell would ‘know’ which simple cells it should connect to, *i.e.*, which simple cells do represent the same object at different locations? Note that a standard Hebbian rule, that learns conjunctions of inputs, does not work here, as only one (or a few) of the targeted simple cells will be activated at once. Instead, a learning rule is needed to learn *disjunctions* of inputs.

Several authors have proposed to use temporal continuity to learn complex cells from transformation sequences (Perrett et al., 1984; Földiák, 1991; Hietanen et al., 1992; Wallis et al., 1993; Wachsmuth et al., 1994; Wallis and Rolls, 1997; Rolls and Milward, 2000; Wiskott and Sejnowski, 2002). This can be done using an associative learning rules that incorporate a temporal trace of activity in the post-synaptic neuron (Földiák, 1991), exploiting the fact that objects seldom appear or disappear, but are often translated in the visual field. Hence simple units that are activated in close temporal proximity are likely to represent the same object, presumably at different locations. Földiák (1991) proposed a modified Hebbian rule, known as the ‘trace rule’ which constrain synapses to be reinforced when strong inputs coincides with strong average past activity (instead of strong current activity in case of a standard Hebbian rule). This proposal has formed the basis of a large number of algorithms for learning invariances from sequences of images (Becker and Hinton, 1992; Stone and Bray, 1995; Wallis and Rolls, 1997; Bartlett

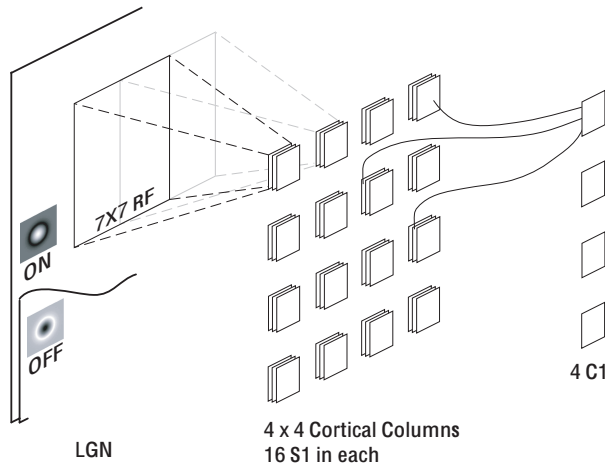


Figure 3: Overview of the specific implementation of the Hubel & Wiesel V1 model used. LGN-like ON- and OFF-center units are modeled by Difference-of-Gaussian (DoG) filters. Simple units (denoted S_1) sample their inputs from a 7×7 grid of LGN-type afferent units. Simple S_1 units are organized in cortical hypercolumns (4×4 grid, 3 pixels apart, 16 S_1 units per hypercolumn). At the next stage, 4 complex units C_1 cells receive inputs from these $4 \times 4 \times 16$ S_1 cells. This paper focuses on the learning of the S_1 to C_1 connectivity.

and Sejnowski, 1998; Stringer and Rolls, 2000; Rolls and Milward, 2000; Wiskott and Sejnowski, 2002; Einhäuser et al., 2002; Spratling, 2005).

However, as pointed out by Spratling (2005), the trace rule by itself is inappropriate when multiple objects are present in a scene: it cannot distinguish which input corresponds to which object, and it may end-up combining multiple objects in the same representation. Hence most trace-rule based algorithm require stimuli to be presented in isolation (Földiák, 1991; Oram and Földiák, 1996; Wallis, 1996; Stringer and Rolls, 2000), and would fail to learn from cluttered natural input sequences.

To solve this problem, Spratling made the hypothesis that the same object could not activate two distinct inputs, hence co-active units necessarily correspond to distinct objects. He proposed a learning rule that can exploit this information, and successfully applied it on drifting bar sequences (Spratling, 2005).

However the ‘one object activates one input’ hypothesis is a strong one. It seems incompatible with the redundancy observed in the mammalian brain and reproduced in our model. Instead we propose another hypothesis: from one frame to another the most active inputs are likely to represent the same object. If the hypothesis is true, by restraining the reinforcement to the most active inputs we usually avoid to combine different objects in the same representation (note that this idea was already present in (Einhäuser et al., 2002), although not formulated in those terms).

In this work we focus on the learning of simple S_1

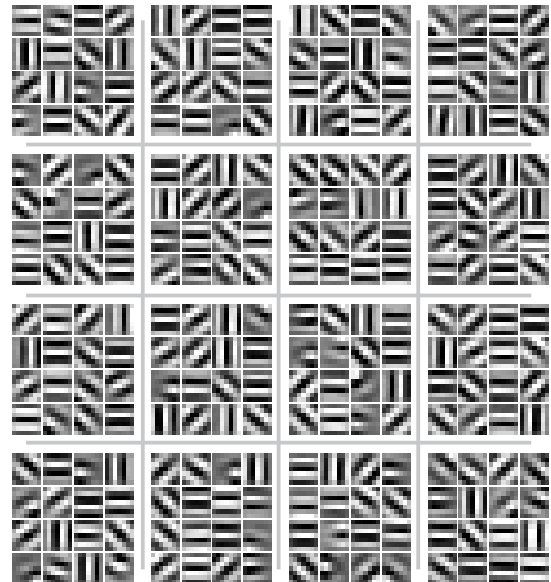


Figure 4: Reconstructed S_1 preferred stimuli for each one of the 4×4 cortical hypercolumns (on this figure the position of the reconstructions within a cortical column is arbitrary). Most units show a Gabor-like selectivity similar to what has been previously reported in the literature (see text).

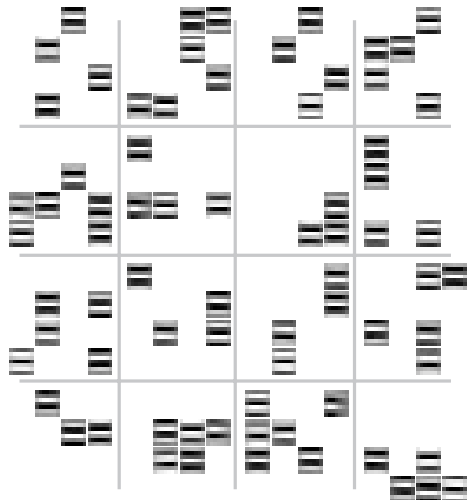
and complex C_1 units (see Fig. 3), which constitutes a direct implementation of the Hubel and Wiesel (1962) model of striate cortex (see Box 1). The goal of a C_1 unit is to pool over S_1 units with the same preferred orientation, but with shifted receptive fields. In this context our hypothesis becomes: ‘in a given neighborhood, the dominant orientation is likely to be the same from one frame to another’. As our results suggests (see later), this constitutes a reasonable hypothesis, which leads to appropriate pooling.

2 Results

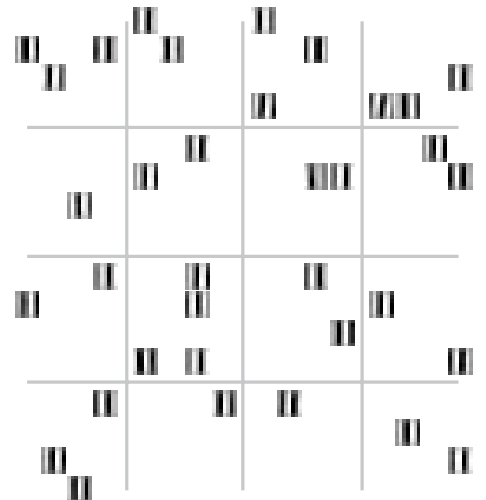
We tested the proposed learning mechanisms in a 3 layer feedforward network mimicking the Lateral Geniculate Nucleus (LGN) and V1 (see Fig. 3). Details of the implementation can be found in Section 4.

The stimuli we used were provided by Betsch et al. (2004). The videos were captured by CCD cameras attached to a cat’s head, while the animal was exploring several outdoor environments. These videos approximate the input to which the visual system is naturally exposed, although eye movements are not taken into account.

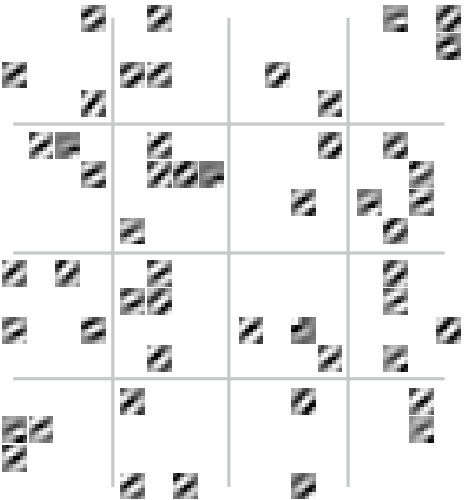
To simplify the computations, learning was done in two phases: First S_1 units learned their selectivity through competitive Hebbian learning. After convergence, plasticity at the S_1 stage was switched off and learning at the complex C_1 unit level started. In a more realistic scenario, this two-phase learning scheme could be approximated with a slow time constant for learning at the S_1 stage and a faster time constant at the C_1 stage.



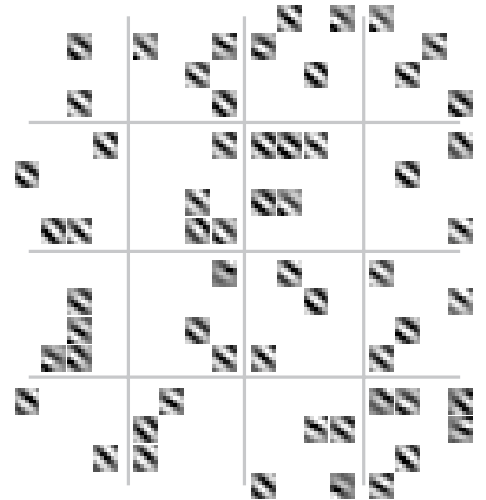
(a) S_1 units ($n=73$) that remain connected to C_1 unit # 1 after learning



(b) S_1 units ($n=35$) that remain connected to C_1 unit # 2 after learning



(c) S_1 units ($n=59$) that remain connected to C_1 unit # 3 after learning



(d) S_1 units ($n=38$) that remain connected to C_1 unit # 4 after learning

Figure 5: Pools of S_1 units connected to each C_1 unit. For *e.g.*, C_1 unit # 1 became selective for horizontal bars: After learning only 73 S_1 units (out of 256) remain connected to the C_1 unit, and they are all tuned to an horizontal bar, but at different positions (corresponding to different cortical columns; on this figure the positions of the reconstructions correspond to their positions in Fig. 4).

2.1 Simple cells

After about 9 hours of simulated time S_1 units have learned a Gabor-like selectivity (see Fig. 4) similar to what has been previously reported for cortical cells (Hubel and Wiesel, 1959, 1962, 1965, 1968; Schiller et al., 1976a,b,c; DeValois et al., 1982a,b; Jones and Palmer, 1987; Ringach, 2002). In particular, receptive fields are localized, tuned to specific spatial frequencies in a given orientation. In this experiment, only four dominant orientations emerged spanning the full range of orientations with 45° increment: 0° , 45° , 90° and 135° . Interestingly, in an another experiment using S_1 receptive

fields larger than the 7×7 receptive field sizes used here, we found instead a continuum of orientations. The fact that we obtain only four orientations here is likely to be a discretization artifact. With this caveat in mind, in the following we used the 7×7 RF sizes (see Table 1), which match the receptive field sizes of cat LGN cells.

Our results are in line with previous studies that have shown that competitive Hebbian learning with DoG inputs leads to Gabor-like selectivity (see for instance (DeLorme et al., 2001; Einhäuser et al., 2002; Guyonneau, 2006) and (Olshausen and Field, 1996) for a more sophisticated model).

2.2 Complex cells

In phase 2, to learn the receptive fields of the C_1 units, we turned off learning at the S_1 stage and began to learn the $S_1 - C_1$ connectivity. This was done using a learning rule that reinforce the synapse between the currently most activated S_1 unit and the previously most activated C_1 unit (see Section 4.4). After 19 hours of simulated time, we ended up with binary $S_1 - C_1$ weights, and each C_1 remained connected to a pool of S_1 with the same preferred orientation, eventually in different cortical columns (see Fig. 5). Hence by taking the (soft) maximum response among its pool, a C_1 unit becomes shift-invariant and inherits its orientation selectivity from its input S_1 units.

In total 38 S_1 units were not selected by any C_1 (see Fig. 6). They either had an atypical preferred stimulus or were tuned to an horizontal bar, which, because of a possible bias in the training data (maybe due to horizontal head movements), is over-represented at the S_1 level. In addition we did not find any S_1 unit selected by more than one C_1 unit. In other words, the pools Fig. 5(a), 5(b), 5(c), 5(d) and 6 were all disjoint. Note that superimposing those 5 figures leads to Fig. 4.

We also experimented with other learning rules within the same architecture. We reimplemented Földiák’s original trace rule (Földiák, 1991) (see Section 4, Eq. 12). As expected, the learning rule failed mainly due to the fact that input frames do not contain isolated edges but instead edges with multiple orientations. This, in turn, leads to complex units that pool over multiple orientations.

We also implemented Spratling’s learning rule (Spratling, 2005), which failed in a similar way, because the hypothesis that ‘one edge activates one S_1 unit’ is violated here.

Finally we re-implemented the rule by Einhäuser et al. (2002) (see Section 4, Eq. 10). We reproduced their main results and the learning rule generated a continuum of $S_1 - C_1$ weights (as opposed to our binary weights). The strongest synapses of a given complex cell did correspond to simple cells with the same preferred orientation but, undesirably, the complex cells had also formed connection to other simple cells with distinct preferred orientation (see also Section 3).

3 Discussion

Contrary to most previous approaches (Földiák, 1990, 1991; Wallis and Rolls, 1997; Stringer and Rolls, 2000; Rolls and Milward, 2000; Spratling, 2005), our approach deals with natural image sequences, as opposed to artificial stimuli such as drifting bars. For any given algorithm to be plausible a necessary condition (although not sufficient) is that it can handle natural images, which brings supplementary difficulties such as noise, clutter and absence of relevant stimuli. Models

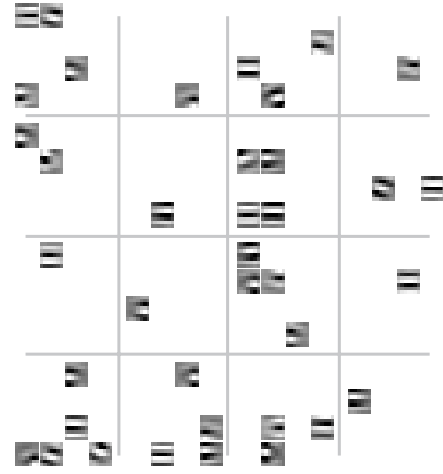


Figure 6: The 38 S_1 cells that were not connected to any C_1 .

that process simpler stimuli may be useful to illustrate a given mechanism, but the ultimate goal should be to deal with natural images, just like humans do. To our knowledge, the only model for the learning of simple and complex cells, which has been shown to work on natural image sequences is the one by Einhäuser et al. (2002). Our work extends the study by Einhäuser et al. (2002) in several significant ways.

First, by using soft-bounds in the weight update rule (see Equation 9), the proposed algorithm converges towards input weights to a complex unit that are binary. This means that each complex unit is strongly connected to a pool of simple units that all have the same preferred orientation. This leads to complex units with an orientation bandwidth similar to the orientation bandwidth of simple units (see (Serre and Riesenhuber, 2004)) in agreement with experimental data (DeValois et al., 1982b). Conversely, the algorithm by Einhäuser *et al.* generates a continuum of synaptic weights and although weaker, some of the connections to simple units to non-preferred orientations remained thus broadening the orientation bandwidth from simple to complex units.

Another important difference with the learning rule used in (Einhäuser et al., 2002) is that our modified Hebbian learning rule is based on the correlation between the current inputs to a complex unit and its output at the previous time step (as opposed to previous input and current output in (Einhäuser et al., 2002)). This was suggested in (Rolls and Milward, 2000). Here we found empirically that it leads to faster and more robust learning. It also turns out to be easier to implement in biophysical circuits: Because of synaptic delays, it is in fact very natural to consider the current input to a unit and its output to the previous frames a few tens of milliseconds earlier. Measuring correlations between past inputs and current output would need an additional mechanism to store the current input for future use.

Finally our approach tends to be simpler than most of the previous ones. The inputs to the model are raw gray-level images without any pre-processing such as low pass filtering or whitening. Also the proposed algorithm does not require any weight normalization and all the learning rules used are local.

Our neurophysiologically-plausible approach also contrasts with objective function approaches, which optimize a given function (such as sparseness (Olshausen and Field, 1996; Rehn and Sommer, 2007) (minimizing the number of units active for any input), statistical independence (Bell and Sejnowski, 1997; van Hateren and Ruderman, 1998; van Hateren and van der Schaaf, 1998; Hyvärinen and Hoyer, 2001) or even temporal continuity and slowness (Wiskott and Sejnowski, 2002; Körding et al., 2004; Berkes and Wiskott, 2005) in a non-biologically plausible way. Such normative model can provide insights as to why receptive fields look the way they do. Indeed such models have made quantitative predictions, which have been compared to neural data (see (van Hateren and Ruderman, 1998; van Hateren and van der Schaaf, 1998; Ringach, 2002) for instance). However, such approaches ignore the computational constraints imposed by the environment and are agnostic about how such learning could be implemented in the cortex.

Fortunately some of them lead to reasonable rules (e.g., (Olshausen and Field, 1996; Sprekeler et al., 2007)) and connections can be drawn between the two classes of approaches. For example, Sprekeler *et al.* recently showed that Slow Feature Analysis (SFA) is in fact equivalent to the trace rule, and could be implemented by Spike Timing Dependant Plasticity (Sprekeler et al., 2007).

Finally our approach constitutes a plausibility proofs for most models of the visual cortex (Fukushima, 1980; Riesenhuber and Poggio, 1999; Ullman et al., 2002; Serre et al., 2007; Masquelier and Thorpe, 2007), which typically learn the tuning of units at one location and simply ‘replicate’ the tuning of units at all locations. This is not the approach we undertook in this work: The 4×4 grid of S_1 units (16 at each location) are all learned independently and indeed are not identical. We then suggested a mechanism to pool together cells with similar preferred stimulus. The success of our approach validates the simplifying assumption of weight-sharing. However we still have to test the proposed mechanisms for higher order neurons.

The idea of exploiting temporal continuity to build invariant representations finds partial support from psychophysical studies, which have suggested that human observers tend to associate together successively presented views of paperclip objects (Sinha and Poggio, 1996) or faces (Wallis and Bülthoff, 2001). The idea also seems consistent – as pointed out by Stryker (Stryker, 1991; Földiák, 1998; Giese and Poggio, 2003) –



Figure 7: Videos: the world from a cat’s perspective (Betsch et al., 2004).

with an electrophysiological study by Miyashita (1988), who showed, that training a monkey with a fixed sequence of image patterns lead to a correlated activity between those same patterns during the delayed activity.

Finally this class of algorithms lead to an interesting prediction made by Einhäuser et al. (2002), namely that the selectivity of complex units could be impaired by rearing an animal in an environment in which temporal continuity would be disrupted (for instance using a stroboscopic light or constantly flashing uncorrelated pictures). We verified this prediction on our model and found that randomly shuffling the frames of the videos had no impact on the development of the simple S_1 units while the selectivity of the complex C_1 units was significantly impaired (all the synapses between the simple and complex units ended up depressed).

To conclude, although this study could be pushed further – in particular the proposed mechanism should be implemented on spiking neurons, and should be tested on higher order neurons – it does constitute a plausibility proof that invariances could be learned using a simple trace-rule, even in natural cluttered environment.

4 Methods

4.1 Stimuli: the world from a cat’s perspective

The videos used were taken from (Betsch et al., 2004) (see Fig. 7). The camera spans a visual angle of 71° by 53° and its resolution is 320×240 pixels. Hence each pixel corresponds to about 13 min of arc. We only used the first six videos (from thirteen total) for a total duration of about 11 minutes. Spatio-temporal patches were extracted from these videos at fixed points from a 9×11 grid (sampled every 25 pixels). These 99 se-

quences were concatenated leading to a total of about 19 hours of video (about 1.6 million frames).

In the following, we set the receptive field sizes for model LGN-like, simple S_1 and complex C_1 units to the average values reported in the literature for foveal cells in the cat visual cortex (Hubel and Wiesel, 1968). We did not model the increase in RF size with eccentricity and assumed that foveal values stood everywhere. This leads to receptive field sizes for the three layers that are summarized in Table 1.

Table 1: Receptive field sizes in pixels, and in degree of visual angle.

	ON – OFF	S_1	C_1
Pixels	7	13	22
Degrees	1.6	2.9	4.9

4.2 LGN ON- and OFF-center unit layer

Gray level images are first analyzed by an array of LGN-like units that correspond to 7×7 Difference-of-Gaussian (DoG) filters:

$$DoG = \frac{1}{2\pi} \left(\frac{1}{\sigma_1} e^{-\frac{r^2}{2\sigma_1^2}} - \frac{1}{\sigma_2} e^{-\frac{r^2}{2\sigma_2^2}} \right) \quad (1)$$

We used $\sigma_2 = 1.4$ and $\sigma_2/\sigma_1 = 1.6$ to make the DOG receptive fields approximate a Laplacian filter profile, which in turn resembles the receptive fields of biological retinal ganglion cells (Marr and Hildreth, 1980). Positive values ended in the ON-center cell map, and the absolute value of negative values in the OFF-center cell map.

4.3 S_1 layer: competitive Hebbian learning

Model S_1 units are organized on a 4×4 grid of cortical columns. Each column contains 16 S_1 units (see Fig. 3). The distance between columns was set to 3 pixels (*i.e.*, about half a degree of visual angle). Each S_1 unit received their inputs from a 7×7 grid of afferent LGN-like units (both ON- and OFF-center) for a total of $7 \times 7 \times 2$ input units. S_1 units perform a bell-shape TUNING (see (Serre et al., 2005a) for details) function which can be approximated by the following static mathematical operation (see Box 2):

$$y_{\text{raw}} = \frac{\sum_{j=1}^n w_j \cdot x_j^p}{k + (\sum_{j=1}^n x_j^q)^r} \quad (2)$$

Here for the S_1 cells we set the parameters to: $k = 0$, $p = 1$, $q = 2$ and $r = 1/2$, which is exactly a normalized dot-product:

$$y_{\text{raw}} = \frac{w \cdot x}{\|x\|} \quad (3)$$

The reader should refer to (Knoblich et al., 2007) for biophysical circuits of integrate and fire neurons that use realistic parameters of synaptic transmission approximating Eq. 3.

The response of a simple S_1 unit is maximal if the input vector x is collinear to the synaptic weight vector w (*i.e.*, the preferred stimulus of the unit). As the pattern of input becomes more dissimilar to the preferred stimulus, the response of the unit decreases monotonically in a bell-shape-like way (*i.e.*, the cosine of the angle between the two vectors).

The unit activity y_{raw} is further normalized by the recent unit history, *i.e.*, a ‘running average (denoted by $tr(\cdot)$) of the raw activities over past few frames’:

$$y = \frac{y_{\text{raw}}}{tr(y_{\text{raw}})} \quad (4)$$

Such unit history is often referred to as a (memory) *trace* (Földiák, 1991; Wallis, 1996; Wallis and Rolls, 1997; Stringer and Rolls, 2000; Rolls and Milward, 2000). For our model S_1 unit, such normalization by the trace approximates adaptation effects. One can think of y_{raw} as the membrane potential of the unit while y approximates the instantaneous firing rate of the unit over short time intervals: units that have been strongly active will become less responsive. While non-critical, this normalization by the trace significantly speeds-up the convergence of the learning algorithm by balancing the activity between all S_1 units (the response of units with a record of high recent activity is reduced while the response of units which have not been active in the recent past is enhanced).³

The initial w weights of all the S_1 units were initialized at random (sampled from a uniform distribution on the $[0,1]$ interval). In each cortical hypercolumn only the most active cell is allowed to fire (1-Winner-Take-All mechanism). However, it will do so if and only if its activity reaches its threshold T . It will then trigger the (modified) Hebbian rule:

$$\Delta w = \alpha \cdot y \cdot (x - w) \quad (5)$$

The $-w$, added to the standard Hebb rule, allows to keep the w bounded. However, the learning rule is still fully local.

The winner then updates its threshold as follows:

$$T = y \quad (6)$$

At each time step, all thresholds are decreased as follows:

$$T = (1 - \eta) \cdot T \quad (7)$$

There is experimental evidence for such threshold modulations in pyramidal neurons, which contribute to homeostatic regulation of firing rates. For example, Desai *et al.* showed that depriving neurons of activity for

two days increased sensitivity to current injection (Desai et al., 1999).

At each time step the traces are updated as follows:

$$tr(y_{\text{raw}}) = \frac{y_{\text{raw}}}{\nu} + \left(1 - \frac{1}{\nu}\right) \cdot tr(y_{\text{raw}}) \quad (8)$$

We used $\eta = 2^{-15}$ and $\nu = 100$. It was found useful to geometrically increase the learning rate α for each S_1 cell every 10 weight updates, starting from an initial value of 0.01 and ending at 0.1 after 200 weight updates. Only half of the 1,683,891 frames were needed to reach convergence.

4.4 C_1 Layer: pool together consecutive winners

4 C_1 cells receive inputs from the $4 \times 4 \times 16$ S_1 cells through synapses with weight $w \in [0, 1]$ (initially set to .75).

Each C_1 cell's activity is computed using Eq. 2, but this time with $p = 6$ (and still $q = 2$ and $r = 1/2$). It has been shown that such operation performs a SOFT-MAX (Yu et al., 2002), and biophysical circuits to implement it have been proposed in (Knoblich et al., 2007).

Winner-Take-All mechanisms select the C_1 winner at time $t - \Delta t$ (previous frame), $J_{t-\Delta t}$, and the current S_1 winner at time t (current frame), I_t . The synapse between them is reinforced, while all the other synapses of $J_{t-\Delta t}$ are depressed:

$$\Delta w_{iJ_{t-\Delta t}} = \begin{cases} a^+ \cdot w_{iJ_{t-\Delta t}} \cdot (1 - w_{iJ_{t-\Delta t}}) & \text{if } i = I_t \\ a^- \cdot w_{iJ_{t-\Delta t}} \cdot (1 - w_{iJ_{t-\Delta t}}) & \text{otherwise.} \end{cases} \quad (9)$$

Synaptic weights for the non-winning C_1 cells are unchanged.

This learning rule was inspired by previous work on Spike Timing Dependent Plasticity (STDP) (Masquelier and Thorpe, 2007). The multiplicative term $w_{iJ_{t-\Delta t}} \cdot (1 - w_{iJ_{t-\Delta t}})$ ensures the weight remains in the range [0,1] (excitatory synapses) and implements a soft bound effect: when the weight approaches a bound, weight changes tend toward zero, while the most plastic synapses are those in an intermediate state.

As recommended by (Rolls and Milward, 2000) we chose to exploit correlations between the previous output and the current input (as opposed to current output and previous output, as Einhuser et al. (2002)). We empirically confirmed that learning was indeed more robust this way.

It was found useful to geometrically increase the learning rates every 1000 iterations, while maintaining the a^+/a^- ratio at a constant value (-170). We started with $a^+ = 2^{-3}$ and set the increase factor so as to reach $a^+ = 2^{-1}$ at the end of the simulation.

4.5 Main differences with Einhuser et al. 2002

- Learning rule for complex cells: first Einhuser et al. select the C_1 winner at time t (current frame), J_t , and the previous S_1 winner at time $t - \Delta t$ (previous frame), $I_{t-\Delta t}$. The synapse between them is reinforced, while all the other synapses of J_t are depressed. Second, Einhuser et al. (2002) use a different weight update rule:

$$\Delta w_{iJ_t} = \begin{cases} \alpha \cdot (1 - w_{iJ_t}) & \text{if } i = I_{t-\Delta t} \\ -\alpha \cdot w_{iJ_t} & \text{otherwise.} \end{cases} \quad (10)$$

This learning rule leads to a continuum of weights at the end (as opposed to binary weights). Tests have shown that the problem persists if (like us) we select the C_1 winner at time $t - \Delta t$ (previous frame), $J_{t-\Delta t}$, and the current S_1 winner at time t (current frame), I_t , and apply the rule by Einhuser et al. (2002):

$$\Delta w_{iJ_{t-\Delta t}} = \begin{cases} \alpha \cdot (1 - w_{iJ_{t-\Delta t}}) & \text{if } i = I_t \\ -\alpha \cdot w_{iJ_{t-\Delta t}} & \text{otherwise} \end{cases} \quad (11)$$

This suggests that the problem comes from the weight update rule that was used in (Einhuser et al., 2002), and not from the type of correlation involved.

- In the model by (Einhuser et al., 2002) the activity of the units is normalized by the “trace” (see earlier) also at the complex cell level.
- The model by (Einhuser et al., 2002) is similar to an energy-type model of complex cells such that simple and complex cells have identical receptive field sizes. In particular, the model does not account for the increase in RF sizes between S_1 and C_1 units (typically doubling (Hubel and Wiesel, 1962)).

4.6 Foldiak’s original trace rule

We also reimplemented Foldiak’s original trace rule (Foldiak, 1991) (see Section 2), which is given by:

$$\Delta w = \alpha \cdot tr(y) \cdot (x - w). \quad (12)$$

Notes

¹As suggested by several authors (Foldiak and Young, 1995; Perrett et al., 1998; Shadlen and Newsome, 1998; Keysers et al., 2001; Serre et al., 2005a), because of the strong temporal constraints imposed on the cortical circuits (*i.e.*, computations at each stage have to be performed within very small temporal windows of 10 – 30 ms (Thorpe and Imbert, 1989; Thorpe and Fabre-Thorpe, 2001; Keysers

et al., 2001; Rolls, 2004) under which single neurons can transmit only very few spikes), the basic units of processing are likely to be modules of hundreds of neurons with similar selectivities rather than individual cells. Such computational modules could be implemented by cortical columns (Mountcastle, 1957, 1997). To paraphrase Mountcastle (Mountcastle, 1997): “the effective unit of operation in such a distributed system is not the single neuron and its axon, but groups of cells with similar functional properties and anatomical connections”.

²For the past two decades several studies (in V1 for the most part) have provided evidence for the involvement of GABAergic circuits in shaping the response of neurons (Sillito, 1984; Douglas and Martin, 1991; Ferster and Miller, 2000). Direct evidence for the existence of divisive inhibition comes from an intracellular recording study in V1 (Borg-Graham and Fregnac, 1998). Wilson et al. (1994) also showed the existence of neighboring pairs of pyramidal cells / fast-spiking interneurons (presumably inhibitory) in the prefrontal cortex with inverted responses (*i.e.*, phased excitatory/inhibitory responses). The pyramidal cell could provide the substrate for the weighted sum while the fast-spiking neuron would provide the normalization term.

³Empirically we found that the learning algorithm would still converge without the normalization term. However the distribution of preferred orientations among the learned S_1 units would be far less balanced than in the full learning algorithm (the number of horizontal units would outweigh the number of vertical ones).

Acknowledgments

We would like to thank W. Einhäuser for sharing with us with the catcam video stimuli used in this work.

References

- Amit, Y. and Mascaró, M. (2003). An integrated network for invariant visual detection and recognition. *Vis. Res.*, 43(19):2073–2088.
- Atick, J. (1992). Could information theory provide an ecological theory of sensory processing. *Network: Computation in Neural Systems*, 3:213–251.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.*, 61:183–193.
- Baker, C., Behrmann, M., and Olson, C. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat. Neurosci.*, 5:1210–1216.
- Barlow, H. (1961). *Sensory Communication*, chapter Possible principles underlying the transformation of sensory messages, pages 217–234. MIT Press, Cambridge, MA, wa rosenblith edition.
- Bartlett, M. S. and Sejnowski, T. J. (1998). Learning viewpoint-invariant face representations from visual experience in an attractor network. *Network*, 9(3):399–417.
- Becker, S. and Hinton, G. (1992). A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355:161–163.
- Bell, A. J. and Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Res*, 37(23):3327–3338.
- Berkes, P. and Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision*, 5(6):579–602.
- Berzhanskaya, J., Grossberg, S., and Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spat Vis*, 20(4):337–395.
- Betsch, B., Einhäuser, W., Körding, K., and König, P. (2004). The world from a cat’s perspective - statistics of natural videos. *Biological Cybernetics*, 90(1):41–50.
- Booth, M. C. and Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex*, 8(6):510–523.
- Borg-Graham, L. and Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393:369–373.
- Callaway, E. (1998). Visual scenes and cortical neurons: What you see is what you get. *Proc. Nat. Acad. Sci. USA*, 95(7):3344–3345.
- Carandini, M. and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264:1333–1336.
- Coppola, D., Purves, H., McCoy, A., and Purves, D. (1998). The distribution of oriented contours in the real world. *Proc. Nat. Acad. Sci. USA*, 95(7):4002–4006.
- Crist, R. E., Li, W., and Gilbert, C. D. (2001). Learning to see: experience and attention in primary visual cortex. *Nat. Neurosci.*, 4:519–525.
- DeAngelis, G., Anzai, A., Ohzawa, I., and Freeman, R. (1995). Receptive field structure in the visual cortex: Does selective stimulation induce plasticity? *Proc. Nat. Acad. Sci. USA*, 92:9682–9686.
- Deco, G. and Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res*, 44(6):621–42.
- Deco, G. and Rolls, E. T. (2005). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *J Neurophysiol*, 94(1):295–313.
- Delorme, A., Perrinet, L., Thorpe, S., and M., S. (2001). Networks of integrate-and-fire neurons using rank order coding B: Spike timing dependent plasticity and emergence of orientation selectivity. *Neurocomputing*, 38-40:539–545.
- Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nat. Neurosci.*, 2:515–520.
- DeValois, R., Albrecht, D., and Thorell, L. (1982a). Spatial frequency selectivity of cells in the macaque visual cortex. *Vision Research*, 22:545–559.

- DeValois, R., Yund, E., and Hepler, N. (1982b). The orientation and direction selectivity of cells in macaque visual cortex. *Vis. Res.*, 22:531–544.
- Dolan, R., Fink, G., Rolls, E., Booth, M., Holmes, A., Frackowiak, R., and Friston, K. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, 389(6651):596–599.
- Douglas, R. J. and Martin, K. A. (1991). A functional microcircuit for cat visual cortex. *J. Physiol. (Lond.)*, 440:735–69.
- Einhäuser, W., Kayser, C., König, P., and Körding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur J Neurosci*, 15(3):475–486.
- Erickson, C. A., Jagadeesh, B., and Desimone, R. (2000). Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nat. Neurosci.*, 3:1143–1148.
- Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, 23:441–471.
- Finn, I. and Ferster, D. (2007). Computational diversity in complex cells of cat primary visual cortex. *J. Neurosci.*, 27(36):9638–48.
- Földiák, P. (1990). Forming sparse representations by local anti-hebbian learning. *Biol Cybern*, 64(2):165–170.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3:194–200.
- Földiák, P. (1998). Learning constancies for object perception. In Walsh, V. and Kulikowski, J. J., editors, *Perceptual Constancy: Why things look as they do*, pages 144–172. Cambridge Univ. Press, Cambridge, UK.
- Földiák, P. and Young, M. (1995). *The handbook of brain theory and neural networks*, chapter Sparse coding in the primate cortex, pages 895–898. MIT Press, Cambridge.
- Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci*, 23(12):5235–5246.
- Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex*, 16(11):1631–1644.
- Frégnac, Y., Shulz, D., Thorpe, S., and Bienenstock, E. (1988). A cellular analogue of visual cortical plasticity. *Nature*, 333(6171):367–370.
- Frégnac, Y. and Shulz, D. E. (1999). Activity-dependent regulation of receptive field properties of cat area 17 by supervised hebbian learning. *J Neurobiol*, 41(1):69–82.
- Fukushima, K. (1980). Neocognitron : a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*, 36(4):193–202.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nat Neurosci*, 2(6):568–73.
- Gawne, T. J. and Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J. Neurophys.*, 88:1128–1135.
- Ghose, G. M. (2004). Learning in mammalian sensory cortex. *Curr Opin Neurobiol*, 14(4):513–518.
- Ghose, G. M., Yang, T., and Maunsell, J. H. R. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophys.*, 87:1867–1888.
- Giese, M. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements and action. *Nature Reviews Neuroscience*, 4:179–192.
- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:213–257.
- Guyonneau, R. (2006). *Codage par latence et STDP: des stratégies temporelles pour expliquer le traitement visuel rapide*. PhD thesis, Université Toulouse III - Paul Sabatier.
- Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J., and Ditttrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp. Brain Res.*, 89:157–171.
- Hochstein, S. and Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36:791–804.
- Horng, S. H. and Sur, M. (2006). Visual activity and cortical rewiring: activity-dependent plasticity of cortical networks. *Prog Brain Res*, 157:3–11.
- Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol*, 160:106–154.
- Hubel, D. and Wiesel, T. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol*, 28:229–289.
- Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat’s striate visual cortex. *J. Phys.*, 148:574–591.
- Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Phys.*, 195:215–243.
- Hung, C., Kreiman, G., Poggio, T., and DiCarlo, J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866.

- Hyvärinen, A. and Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vis. Res.*, 41(18):2413–2423.
- Jagadeesh, B., Chelazzi, L., Mishkin, M., and Desimone, R. (2001). Learning increases stimulus salience in anterior inferior temporal cortex of the macaque. *J. Neurophys.*, 86:290–303.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., and Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, 53(6):891–903.
- Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1233–1258.
- Karni, A. and Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA*, 88:4966–4970.
- Keysers, C., Xiao, D. K., Földiák, P., and Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.*, 13:90–101.
- Knill, D. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.
- Knoblich, U., Bouvrie, J., and T., P. (2007). Biophysical models of neural computation: Max and tuning circuits. Technical report, CBCL Paper, MIT.
- Kobatake, E., Wang, G., and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophys.*, 80:324–330.
- Körding, K., Kayser, C., Einhäuser, W., and König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli? *J. Neurophys.*, 91(1):206–212.
- Kouh, M. and Poggio, T. (2007). A canonical cortical circuit for gaussian-like and max-like operations. *Neural Comp.*, in press.
- Kourtzi, Z., Betts, L. R., Sarkheil, P., and Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biol*, 3(7):e204.
- Kourtzi, Z. and DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Curr Opin Neurobiol*, 16(2):152–158.
- Lampl, I., Ferster, D., Poggio, T., and Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex. *J Neurophysiol*, 92(5):2704–2713.
- Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.*, 5:552–563.
- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London*, B(207):187–217.
- Masquelier, T. and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput Biol*, 3(2):e31.
- McLean, J. and Palmer, L. A. (1998). Plasticity of neuronal response properties in adult cat striate cortex. *Vis Neurosci*, 15(1):177–196.
- Mel, B. W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comp.*, 9:777–804.
- Meliza, C. D. and Dan, Y. (2006). Receptive-field modification in rat visual cortex induced by paired visual stimulation and single-cell spiking. *Neuron*, 49(2):183–189.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335:817–820.
- Mountcastle, V. (1957). Modality and topographic properties of single neurons of cat’s somatic sensory cortex. *J. Neurophys.*, 20:408–434.
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain.*, 120(Part 4):701–22.
- Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.
- Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., and Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *J Neurosci*, 26(50):13025–13036.
- Oram, M. and Földiák, P. (1996). Learning generalisation and localisation: competition for stimulus type and receptive field. *Neurocomputing*, 11(2-4):297–321.
- Pasupathy, A. and Miller, E. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, 24:873–876.
- Perrett, D., Hietanen, J., Oram, M., and Benson, P. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. Roy. Soc. B*, 335:23–30.
- Perrett, D. and Oram, M. (1993). Neurophysiology of shape processing. *Img. Vis. Comput.*, 11:317–333.
- Perrett, D., Oram, M., and Wachsmuth, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, 67:111–145.

- Perrett, D., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A., and Jeeves, M. (1984). Neurones responsive to faces in the temporal cortex: Studies of functional organisation, sensitivity to identity, and relation to perception. *Human Neurobiology*, 3:197–208.
- Perrinet, L., M., S., and Thorpe, S. (2004). Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit. *Neurocomputing*, 57:125–134.
- Poggio, T. and Bizzi, E. (2004). Generalization in vision and motor control. *Nature*, 431(7010):768–774.
- Poggio, T. and Girosi, F. (1990). Networks for approximation and learning. *Proc. IEEE*, 78(9).
- Poggio, T. and Smale, S. (2003). The mathematics of learning: Dealing with data. *Notices of the American Mathematical Society (AMS)*, 50(5).
- Rainer, G., Lee, H., and Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biol*, 2(2):E44.
- Rainer, G. and Miller, E. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, 27:8–10.
- Rehn, M. and Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J Comput Neurosci*, 22(2):135–146.
- Reichardt, W., Poggio, T., and Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual system of the fly – II: Towards the neural circuitry. *Biol. Cyb.*, 46:1–30.
- Richards, W., Feldman, J., and Jepson, A. (1992). From features to perceptual categories. In *Proc. British Machine Vision Conference*, pages 99–108.
- Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci*, 2(11):1019–1025.
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol*, 88(1):455–463.
- Rolls, E. (1995). Learning mechanisms in the temporal lobe visual cortex. *Behav. Brain Res.*, 66(1-2):177–185.
- Rolls, E. (2004). *Invariant object and face recognition*. In *The Visual Neurosciences*, pages 1165–1178. MIT Press, Cambridge, MA.
- Rolls, E. and Deco, G. (2002). *Computational neuroscience of vision*. Oxford University Press.
- Rolls, E. and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Comput*, 12(11):2547–2572.
- Rolls, E. T. and Stringer, S. M. (2006). Invariant visual object recognition: a model, with lighting invariance. *J Physiol Paris*, 100(1-3):43–62.
- Ruderman, D. (1994). The statistics of natural images. *Network : Computation in Neural Systems*, 5:598–605.
- Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp Brain Res*, 77(1):23–30.
- Schiller, P. H., Finlay, B. L., and Volman, S. F. (1976a). Quantitative studies of single-cell properties in monkey striate cortex I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.*, 39(6):1288–1319.
- Schiller, P. H., Finlay, B. L., and Volman, S. F. (1976b). Quantitative studies of single-cell properties in monkey striate cortex II. Orientation specificity and ocular dominance. *J. Neurophysiol.*, 39(6):1334–51.
- Schiller, P. H., Finlay, B. L., and Volman, S. F. (1976c). Quantitative studies of single-cell properties in monkey striate cortex III. Spatial frequency. *J. Neurophysiol.*, 39(6):1334–1351.
- Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, 412:549–553.
- Schuett, S., Bonhoeffer, T., and Hubener, M. (2001). Pairing-induced changes of orientation maps in cat visual cortex. *Neuron*, 32:325–337.
- Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005a). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *Massachusetts Institute of Technology*, CBCL Paper #259 / AI Memo #2005-036.
- Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Nat. Acad. Sci. USA*, 104(15).
- Serre, T. and Poggio, T. (2005). Standard model v2.0: How visual cortex might learn a universal dictionary of shape-components. *Journal of Vision*, 5(5):742.
- Serre, T. and Riesenhuber, M. (2004). Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. *AI Memo 2004-017 / CBCL Memo 239*, MIT, Cambridge, MA.
- Serre, T., Wolf, L., and Poggio, T. (2005b). Object recognition with features inspired by visual cortex. *CVPR*, 2:994–1000.
- Shadlen, M. and Newsome, W. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.*, 18(10):3870–3896.
- Sigala, N. and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415(6869):318–20.

- Sillito, A. (1984). *Functional properties of cortical cells*, volume 2, chapter Functional considerations of the operation of GABAergic inhibitory processes in the visual cortex, pages 91–117. New York: Plenum Press.
- Simoncelli, E. and Olshausen, B. (2001). Natural image statistics and neural representation. *Ann. Rev. Neurosci.*, 24:1193–1216.
- Singer, W., Tretter, F., and Yinon, U. (1982). Evidence for long-term functional plasticity in the visual cortex of adult cats. *J. Neurophys.*, 324:239–248.
- Sinha, P. and Poggio, T. (1996). The role of learning in 3-D form perception. *Nature*, 384:460–463.
- Smith, E. and Lewicki, M. (2006). Efficient auditory coding. *Nature*.
- Spratling, M. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5).
- Sprekeler, H., Michaelis, C., and Wiskott, L. (2007). Slowness: an objective for spike-timing-dependent plasticity? *PLoS Comput Biol*, 3(6):e112.
- Stone, J. and Bray, A. (1995). A learning rule for extracting spatio-temporal invariances. *Network*, 6(3):1–8.
- Stringer, S. and Rolls, E. (2000). Position invariant recognition in the visual system with cluttered environments. *Neural Netw*, 13(3):305–315.
- Stryker, M. P. (1991). Temporal associations. *Nature*, 354:108–109.
- Sutton, R. and Barto, A. (1981). Towards a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.*, 88:135–170.
- Thorpe, S. and Imbert, M. (1989). Biological constraints on connectionist modelling. In *Connectionism in perspective*, pages 63–92. Amsterdam: Elsevier.
- Thorpe, S. J. and Fabre-Thorpe, M. (2001). Neuroscience. seeking categories in the brain. *Science*, 291(5502):260–263.
- Torre, V. and Poggio, T. (1978). A synaptic mechanism possibly underlying directional selectivity motion. *Proc. of the Royal Society London B*, 202:409–416.
- Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci*, 5(7):682–687.
- van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc Biol Sci*, 265(1412):2315–2320.
- van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Biol Sci*, 265(1394):359–366.
- VanRullen, R., Gautrais, J., Delorme, A., and Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems*, 48(1-3):229–239.
- VanRullen, R., Guyonneau, R., and Thorpe, S. (2005). Spike times make sense. *Trends Neurosci*, 28(1):1–4.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100.
- Wachsmuth, E., Oram, M., and Perrett, D. (1994). Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 4:509–522.
- Wallis, G. (1996). Using spatio-temporal correlations to learn invariant object recognition. *Neural Networks*, 9(9):1513–1519.
- Wallis, G. and Bülthoff, H. (2001). Role of temporal association in establishing recognition memory. *Proc. Nat. Acad. Sci. USA*, 98(8):4800–4804.
- Wallis, G. and Rolls, E. (1997). Invariant face and object recognition in the visual system. *Prog Neurobiol*, 51(2):167–194.
- Wallis, G., Rolls, E., and Földiák, P. (1993). Learning invariant responses to the natural transformations of objects. *International Joint Conference on Neural Networks*, 2:1087–1090.
- Wersing, H. and Koerner, E. (2003). Learning optimized features for hierarchical models of invariant recognition. *Neural Comp.*, 15(7):1559–1588.
- Wilson, F., Scialaidhe, S., and Goldman-Rakic, P. (1994). Functional synergism between putative γ -aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proc. Nat. Acad. Sci. USA*, 91:4009–4013.
- Wiskott, L. and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput*, 14(4):715–770.
- Yang, T. and Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.*, 24:1617–1626.
- Yao, H. and Dan, Y. (2001). Stimulus timing-dependent plasticity in cortical processing of orientation. *Neuron*, 32:315–323.
- Yu, A. J., Giese, M. A., and Poggio, T. (2002). Biophysiological plausible implementations of the maximum operation. *Neural Comp.*, 14(12):2857–2881.

